

THE CHINESE UNIVERSITY OF HONG KONG
Department of Mathematics

Mathematical Modeling Project Team
mathmodel@math.cuhk.edu.hk

Modeling with R Shiny (Clustering)

Last updated: October 1, 2026

1 Introduction to the Clustering R Shiny tool 聚類分析R Shiny 工具簡介

R is a powerful statistical programming language that allows us to perform data analysis, which is useful for mathematical modeling. Even better, we have set up a set of online tools using R Shiny so that you can perform various mathematical modelling tasks without any coding background.

R 是一種功能強大的統計程式語言，讓我們能夠進行數據分析，對數學建模十分有用。更理想的是，我們利用R Shiny 建立了一系列線上工具，讓使用者即使不具備程式設計背景，也能夠執行各種數學建模任務。

In particular, the Clustering R Shiny tool provides an interactive way to explore different clustering methods and visualize the clustering results. Users can upload a two-dimensional dataset and apply different clustering methods, including partitional clustering, hierarchical clustering, density-based clustering, and other clustering methods. The resulting clusters can be visualized directly on the graph.

聚類分析R Shiny 工具提供了一個互動式平台，讓使用者探索不同的聚類分析方法並視覺化聚類結果。使用者可以上傳二維數據集，並應用不同的聚類分析方法，包括劃分式聚類、層次聚類、基於密度的聚類，以及其他聚類方法。聚類結果可以直接在圖表中進行視覺化。

Here is the guideline on how to use our Clustering R Shiny tool.

以下是有關如何使用我們的聚類分析R Shiny 工具的指引。

2 Using the Clustering R Shiny tool 使用聚類分析R Shiny 工具

To use the tool, please visit our main website:

如要使用此工具，可前往我們的主網站：

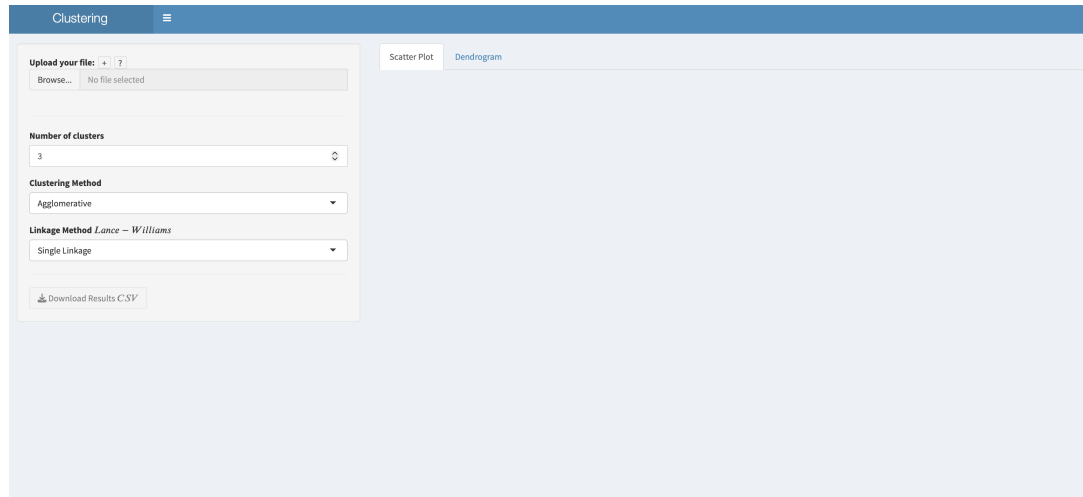
<https://cuhkmathmodel.com/>

for the list of tools, or simply go to
瀏覽工具列表，或直接前往

<https://mathmodelcuhk.shinyapps.io/clustering/>

for the Clustering R Shiny tool.
使用聚類分析R Shiny 工具。

After getting into the website, you can see the following:
進入網站後，你會看到以下畫面：



Don't worry if you think this is complicated! We will guide you step by step, and you can perform clustering using our R Shiny tool.
如果你覺得這個界面有點複雜，不用擔心！我們會逐步引導你，讓你能夠順利使用我們的R Shiny 工具進行聚類分析。

2.1 Data input 數據輸入

The input dataset should contain two columns of data representing the x and y coordinates of the data points.
輸入數據集應包含兩列資料，分別代表數據點的 x 和 y 座標。

Input your data here in csv

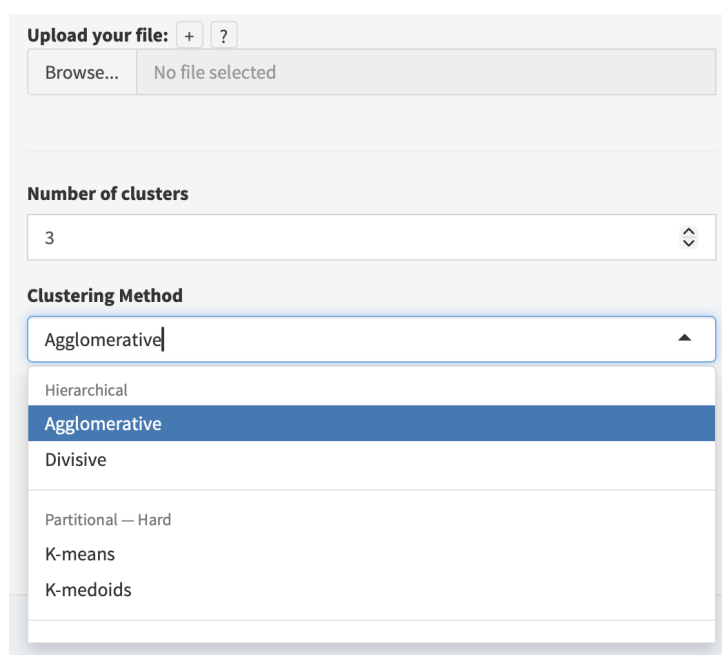
```
x,y
1.5,40.6
3.8,65
8.9,70.8
10,78
12,84
20,108.5
```

The data can be uploaded in CSV, XLSX, or TXT format. Each row represents one data point, with the two columns corresponding to its x and y values.
數據可以以CSV、XLSX 或TXT 格式上傳。每一行代表一個數據點，而兩列資料分別對應其 x 和 y 數值。

2.2 Interactive visualization of different clustering methods 不同聚類分析方法的互動式視覺化

The tool provides several clustering methods for analysing two-dimensional datasets. These methods can be grouped into different categories, including hierarchical clustering, partitional clustering, density-based clustering, and other clustering methods.

工具提供多種聚類分析方法，用於分析二維數據集。這些方法可以分為不同類型，包括層次聚類、劃分式聚類、基於密度的聚類，以及其他聚類方法。



The screenshot shows a web interface for a clustering tool. At the top, there is a section titled "Upload your file:" with a "+" button, a "?" button, and a "Browse..." button. Below this, it says "No file selected". Further down, there is a section titled "Number of clusters" with a dropdown menu currently set to "3". Below that, there is a section titled "Clustering Method" with a dropdown menu. The dropdown menu is open, showing a list of methods: "Agglomerative" (which is highlighted in blue), "Hierarchical", "Divisive", "Partitional — Hard", "K-means", and "K-medoids".

After selecting a clustering method, the corresponding parameters and options will be displayed. Users can adjust the parameters and observe how the clustering result changes in the graph.

選擇聚類分析方法後，介面會顯示相應的參數及選項。使用者可以調整相關參數，並在圖表中觀察聚類結果的變化。

2.3 Hierarchical clustering 層次聚類

Hierarchical clustering builds a hierarchy of clusters and provides a different way of analysing the relationships between data points. The tool supports both agglomerative and divisive hierarchical clustering methods.

層次聚類會建立群組之間的層次結構，從另一個角度分析數據點之間的關係。工具支持凝聚式及分裂式兩種層次聚類方法。

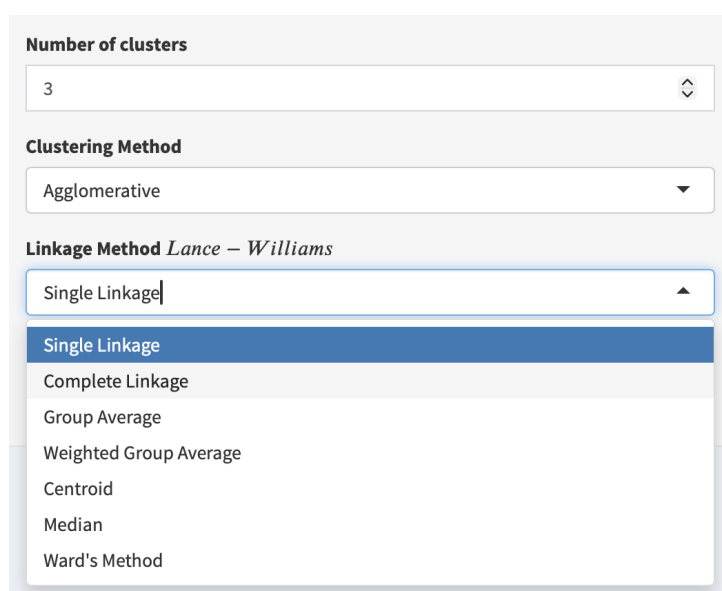
2.3.1 Agglomerative clustering 凝聚式聚類

Agglomerative clustering starts by treating each data point as an individual cluster and progressively merges clusters according to their similarity. The user can specify the desired number of clusters and select a linkage method to determine the distance between clusters. 凝聚式聚類首先將每個數據點視為獨立的群組，然後根據群組之間的相似程度逐步合

併。使用者可以指定所需的群組數目，並選擇連結方法來決定不同群組之間的距離。

After selecting “Agglomerative”, adjust the “Number of clusters” and choose a “Linkage Method”. The available linkage methods include Single Linkage, Complete Linkage, Group Average, Weighted Group Average, Centroid, Median, and Ward’s Method.

選擇「Agglomerative」後，調整「Number of clusters」並選擇「Linkage Method」。可選的連結方法包括單一連結法、完整連結法、群組平均、加權群組平均、形心、中位數及華德最小變異法。



The screenshot displays a user interface for configuring clustering parameters. It includes a 'Number of clusters' input field with the value '3', a 'Clustering Method' dropdown menu currently set to 'Agglomerative', and a 'Linkage Method' dropdown menu that is open, showing a list of options: 'Single Linkage', 'Complete Linkage', 'Group Average', 'Weighted Group Average', 'Centroid', 'Median', and 'Ward's Method'. The 'Linkage Method' dropdown is titled 'Linkage Method Lance – Williams'.

The clustering result will be displayed in the graph. Users can change the number of clusters or linkage method to compare different clustering results.

聚類結果會顯示於圖表中。使用者可以更改群組數目或連結方法，以比較不同的聚類結果。

2.3.2 Divisive clustering 分裂式聚類

Divisive clustering takes the opposite approach. It starts with all data points in one cluster and progressively divides the data into smaller clusters.

分裂式聚類採用相反的方法。它首先將所有數據點視為同一個群組，然後逐步將數據分割成較小的群組。

Select the divisive clustering method and adjust the corresponding parameters. The clustering result will be displayed on the graph for interactive exploration.

選擇分裂式聚類方法並調整相應的參數。聚類結果會顯示於圖表中，方便使用者進行互動式探索。

2.4 Partitional clustering 劃分式聚類

Partitional clustering divides the data points into a specified number of clusters. The Clustering R Shiny tool provides two commonly used partitional clustering methods:

K-means and K-medoids.

劃分式聚類會將數據點劃分為指定數量的群組。聚類分析R Shiny 工具提供兩種常見的劃分式聚類方法：K-means 及K-medoids。

2.4.1 K-means clustering

K-means 聚類

K-means clustering divides the data points into k clusters according to their similarity. The user can choose the number of clusters k , and the tool then assigns the data points to the corresponding clusters.

K-means 聚類根據數據點之間的相似程度，將數據點劃分為 k 個群組。使用者可以選擇群組數目 k ，工具會根據所選的設定將數據點分配到相應的群組。

Adjust the value of k using the corresponding control. The clustering result will be updated accordingly, and the data points belonging to different clusters will be displayed separately on the graph.

使用相應的控制項調整 k 的數值。聚類結果會隨之更新，而屬於不同群組的數據點會在圖表中分別顯示。

2.4.2 K-medoids clustering

K-medoids 聚類

K-medoids is another partitional clustering method. Similar to K-means, the user specifies the number of clusters, and the data points are divided into the selected number of groups. K-medoids 是另一種劃分式聚類方法。與K-means 類似，使用者需要指定群組數目，然後工具會將數據點劃分為所選數量的群組。

Select “K-medoids” as the clustering method and adjust the corresponding parameters. The resulting clusters are displayed interactively on the graph.

選擇「K-medoids」作為聚類分析方法，並調整相應的參數。完成設定後，聚類結果會以互動式圖表顯示。

2.5 Density-based clustering

基於密度的聚類

The tool also provides density-based clustering through the DBSCAN method. Unlike methods that require the number of clusters to be specified directly, DBSCAN identifies clusters according to the density of data points.

工具亦提供基於密度的聚類方法，其中包括DBSCAN。與需要直接指定群組數目的方法不同，DBSCAN 根據數據點的密度來識別不同的群組。

Select “DBSCAN” as the clustering method and adjust the corresponding parameters. The resulting clusters are then displayed on the graph, allowing users to observe groups of data points with different densities and identify points that do not belong to the main clusters.

選擇「DBSCAN」作為聚類分析方法，並調整相應的參數。完成設定後，工具會在圖表中顯示聚類結果，讓使用者觀察不同密度的數據群組，以及識別不屬於主要群組的數據點。

The image shows a user interface for selecting a clustering method. It features a dropdown menu labeled 'Clustering Method' with 'DBSCAN' selected. Below this, there are two input fields: 'Neighborhood radius *eps*' with the value '1' and 'Minimum points *minPts*' with the value '4'. Both input fields have up and down arrow icons on their right sides, indicating they are adjustable.

2.6 Other clustering methods

其他聚類分析方法

In addition to the methods described above, the tool also provides other clustering methods for exploring different structures in two-dimensional datasets. Users can select the corresponding method from the clustering options and adjust the available parameters. 除了上述方法外，工具亦提供其他聚類分析方法，以探索二維數據集中的不同結構。使用者可以從聚類分析選項中選擇相應的方法，並調整可用的參數。

Different clustering methods may produce different grouping results for the same dataset. Therefore, users can compare the visualized results obtained from different methods and investigate how the choice of clustering method affects the identified clusters.

對於同一組數據，不同的聚類分析方法可能會產生不同的分組結果。因此，使用者可以比較不同方法所得到的視覺化結果，並探討聚類方法的選擇如何影響所識別的群組。

2.7 Visualizing the clustering results

聚類結果的視覺化

After selecting a clustering method and adjusting its parameters, the tool displays the clustering result graphically. Since the input data are two-dimensional, the data points can be plotted directly according to their x and y values.

選擇聚類分析方法並調整相關參數後，工具會以圖表形式顯示聚類結果。由於輸入數據為二維數據，因此可以直接根據數據點的 x 和 y 值繪製圖表。

The visualization allows users to identify which data points belong to the same cluster and to observe the spatial structure of the dataset. By changing the clustering method or its parameters, users can interactively compare different clustering results.

透過視覺化結果，使用者可以觀察哪些數據點屬於同一群組，以及數據集在空間上的結構。透過改變聚類分析方法或相關參數，使用者可以互動式地比較不同的聚類結果。

Data points belonging to different clusters are displayed separately, making it easier to distinguish the identified groups.

屬於不同群組的數據點會以不同方式區分顯示，讓使用者更容易辨識所識別出的不同群組。

2.8 Exporting the clustering results

匯出聚類結果

After performing clustering, the tool provides the option to output the cluster labels as a CSV file. The output can be used for further data analysis or mathematical modelling. 完成聚類分析後，工具提供將群組標籤輸出為CSV 檔案的功能。輸出的結果可以用於進一步的數據分析或數學建模。

The cluster label indicates the cluster to which each data point has been assigned. Users can save the output and use it together with the original dataset for subsequent analysis.

群組標籤表示每個數據點所被分配到的群組。使用者可以儲存輸出的結果，並將其與原始數據集結合，以進行後續分析。

3 Conclusion

總結

Clustering is a useful tool in mathematical modelling for identifying groups and patterns in datasets. By considering different clustering methods, including partitional clustering, hierarchical clustering, and density-based clustering, we can explore different structures within two-dimensional datasets. We hope that this R Shiny Clustering tool may help you with your mathematical modelling journey. Good luck!

聚類分析是數學建模中極具價值的工具，可以幫助我們識別數據集中的不同群組及模式。透過考慮不同的聚類分析方法，包括劃分式聚類、層次聚類及基於密度的聚類，我們可以探索二維數據集中的不同結構。我們衷心希望這個R Shiny 聚類分析工具能夠在你的數學建模探索旅程中提供協助。祝你一切順利！